REMARKS

Applicant gratefully acknowledges the examiner's indication that the previous amendment and remarks overcame the anticipation rejections. The new reference and rejections presented in the pending office action are addressed below. The word "for" has been restored to claim 1 as it inadvertently did not appear in the previous claims listing.

Claims 1 and 2 have been rejected under 35 U.S.C. §103(a) as being unpatentable over Matsubayashi in view Van Aelten. A *prima facie* case of obviousness has not been established by the examiner. The motivation to combine cited by the examiner is for Matsubayashi to avoid an explosion of the recognition vocabulary. But Matsubayashi does not use a vocabulary or dictionary. The abstract of Matsubayashi emphasizes in the first sentence that the method described extracts features from a document "without using a word dictionary." Given that Matsubayashi describes a system for performing document searching without the use of a word dictionary, Matsubayashi has no incentive to reduce the size of the non-existent dictionary. Applicant finds no incentive in the prior art for Matsubayashi to consider a lexicon for components of a compound word so as to require a smaller dictionary when in fact Matsubayashi does not use the dictionary at all. The stated motivation to combine is fallacious and fails to support a *prima facie* case of obviousness.

Matsubayashi teaches away from use of a dictionary. Matsubayashi sets forth problems with the use of a word dictionary for word extraction at col. 2, lines 6-32. Given Matsubayashi's indication of problems with the use of dictionaries, this reference clearly teaches away from the use of a lexicon such as that disclosed by Van Aelten. For this additional reason, the cited combination of references is untenable.

Matsubayashi uses calculations to divide up a character string to extract smaller strings for document searching. The frequency of occurrence of the extracted strings in documents in the databases is used in conducting the search to identify relevant documents. While Matsubayashi divides up a character string into n-grams, there is no disclosure, suggestion or teaching of looking up an n-gram in a lexicon. Rather the extracted character strings are used for document searching. Even the portion of

Matsubayashi cited by the examiner for disclosing a word separator (col. 24, lines 19-27) describes document searching rather than making any use of a dictionary. Generally, Matsubayashi describes extracting character strings from an initial document in order to find similar documents in a documents search. The alternative described in column 24, lines 19-27 bases the search on a specific text from which the same process of extracting characteristic strings is used to search for a document.

Van Aelten merely discloses the use of a lexicon with components of compound words rather than all of the compound words themselves. Van Aelten provides no suggestion or incentive to make use of its lexicon in a document searching program that does not make use of any dictionary. There is no incentive to combine the document searching teachings of Matsubayashi with speech recognition of compound words taught by Van Aelten.

Applicant respectfully submits that Matsubayashi's document searching system is not applicable to applicant's invention and that it teaches away from the use of a dictionary or a lexicon. A *prima facie* case of obviousness has not been demonstrated and thus claims 1 and 2 are allowable.

Claims 1 and 10 were rejected under 35 U.S.C. §103(a) as being unpatentable over Iizuka in view of Van Aelten. In accordance with claim 1, after constructing a set of probabilistic breakpoints in a natural language input substrings of the input are traversed to identify linkable components. In accordance with paragraph 86 of Iizuka identified by the examiner, a character string dividing section 205 divides the received document into several words with reference to character joint probabilities. Having divided the input into words the result is output from the system. There is no suggestion, motivation or incentive for further processing involving searching for linkable components. Seeking to interpret compound words forms no part of Iizuka. Indeed, this first embodiment of Iizuka does not make use of a word dictionary. (Iizuka, ¶145). Thus, there is no incentive, suggestion or motivation to combine this teaching of Iizuka with the compound component lexicon of Van Aelten. The examiner suggests that the motivation is to keep the vocabulary from exploding in size, however, given the lack of a dictionary in this embodiment of Iizuka, this suggestion is inapplicable.

Iizuka relates to agglutinative languages such as Japanese, Chinese and some other Asian languages. English and other European languages are non-agglutinative. Agglutinative languages are characterized by a long character string according to which each boundary of neighboring words is not clear. (Iizuka, ¶3). Iizuka notes that in an agglutinative language a compound word may be included in a dictionary. Such a compound word may be further divided into its component parts. (Iizuka, ¶269). Although describing the existence of such compound words, Iizuka provides no incentive for seeking to identify and link the components of the compound word. To the contrary, Iizuka's stated preference in ¶269 was to coarsely divide a character string to identify the compound word as a whole in a dictionary.

Van Aelten on the other hand relates to a different sort of language in which prefixes and suffixes such as "un", "ed", "ing", "able" and "ly" transform a word into a different part of speech. There is no showing that Japanese or the other agglutinative languages described by Iizuka would benefit from linking components as described in Van Aelten. Applicant finds no description in Iizuka of traversing substrings of language input to identify linkable components. Iizuka does not demonstrate any incentive for doing so particularly with respect to the Japanese language.

A third embodiment of Iizuka uses a dictionary in the first step as described at paragraph 226 to examine adjoining characters in a string to determine whether they correspond to any words in a dictionary. In this embodiment, the dictionary is used in place of the use of probabilistic breakpoints. Thus, Iizuka teaches the use of the dictionary in this embodiment to propose a variety of possible words for possible divisions. Character joint probabilities are then used to score the words for determining a most likely pattern. Thus, Iizuka does not disclose traversing substrings of natural-language input delimited by probabilistic breakpoints to identify linkable components.

Claim 1 further refers to interpreting the segmented string as a compound word. While Iizuka describes at paragraph 269 that a compound word is formed by two parts, there is no corresponding discussion of taking any steps to identify or segment the component parts. Thus, Iizuka neither discloses traversing substrings to identify linkable components nor does Iizuka disclose interpreting a segmented string as a compound

word. The examiner further concedes that Iizuka does not disclose locating a linkable component in a lexicon.

Given the vast differences between the agglutinative languages described by Iizuka and the languages for which Van Aelten functions, applicants submits there is no suggestion to use the teachings of one in combination with the other. Van Aelten does not describe the use of probabilistic breakpoints. Iizuka does not disclose traversing substrings of the natural-language input delimited by the probabilistic breakpoints to identify linkable components. Iizuka does not disclose linking components together to interpret a segmented string as a compound word. Applicant respectfully submits that all elements of claim 1 can not be found in Iizuka and Van Aelten. For this reason, the claims should be allowed. Furthermore, there is no suggestion to combine the methods of Iizuka and Van Aelten which work on entirely unrelated languages. Moreover, there is no incentive in Iizuka to seek a smaller lexicon as asserted by the examiner. For these reasons, applicant submits that claim 1 is allowable.

In accordance with claim 10, the traversal of substrings is performed in an order determined by probabilities. Iizuka does not traverse substrings delimited by probabilistic breakpoints to identify linkable components at all. Thus, Iizuka has no need for ordering. Ordering a traversal of substrings is nowhere contemplated by Iizuka. In particular, there is no suggestion, disclosure or teaching of making use of probabilities to determine such an order.

In accordance with the claimed invention, the substrings are traversed so as to identify linkable components by locating each component in a lexicon. The examiner cites the first embodiment of Iizuka at paragraph 144 against claim 10, however, this embodiment does not use a dictionary at all. (Iizuka, ¶145). While Iizuka has discussions of probabilities, there is no suggestion, teaching or disclosure of ordering a traversal of substrings delimited by probabilistic breakpoints to identify linkable components in any order, let alone one determined by probabilities. As such the claimed invention of claim 10 is entirely absent from the cited references.

For all the foregoing reasons, Applicant submits that all claims pending in the present application are patentable over the art of record. Van Aelten has not overcome the deficiencies of Matsubayashi and Iizuka and is not properly combinable with these references in any case. The rejection should be rescinded and a notice of allowance should issue.

Respectfully submitted,

Robert M. Asher
Reg. No. 30,445
Bromberg & Sunstein LLP
125 Summer Street, 11th Floor
Boston, MA 02110-1618
(617) 443-9292
Attorney for Applicant

02639/00A36 539966.1